模式识别-第三次作业

王健

201628015029018

1.

对四个样本采用规范化增广样本表示形式：

四个样本分别为：

$$\begin{aligned} \mathbf{y}_1 &= (1,1,4)^T \\ \mathbf{y}_2 &= (1,2,3)^T \\ \mathbf{y}_3 &= (-1,-4,-1)^T \\ \mathbf{y}_4 &= (-1,-3,-2)^T \end{aligned}$$

根据梯度下降法，有如下更新准则：

$$\begin{aligned} a_{k+1} &= a_k + \eta_k \sum_{\mathbf{y}\in Y_k} \mathbf{y} \\ a_0 &= (0,1,0)^T \\ \eta_k &= 1 \end{aligned}$$

则第一次迭代：

$$Y_k = \{\mathbf{y}_3, \mathbf{y}_4\}$$

$$a_1 = (0,1,0)^T + (-2,-7,-3)^T = (-2,-6,-3)^T$$

第二次迭代：

$$Y_k = \{\mathbf{y}_1, \mathbf{y}_2\}$$

$$a_2 = (-2,-6,-3)^T + (2,3,7)^T = (0,-3,4)^T$$

此时：

$$Y_k = \varnothing$$

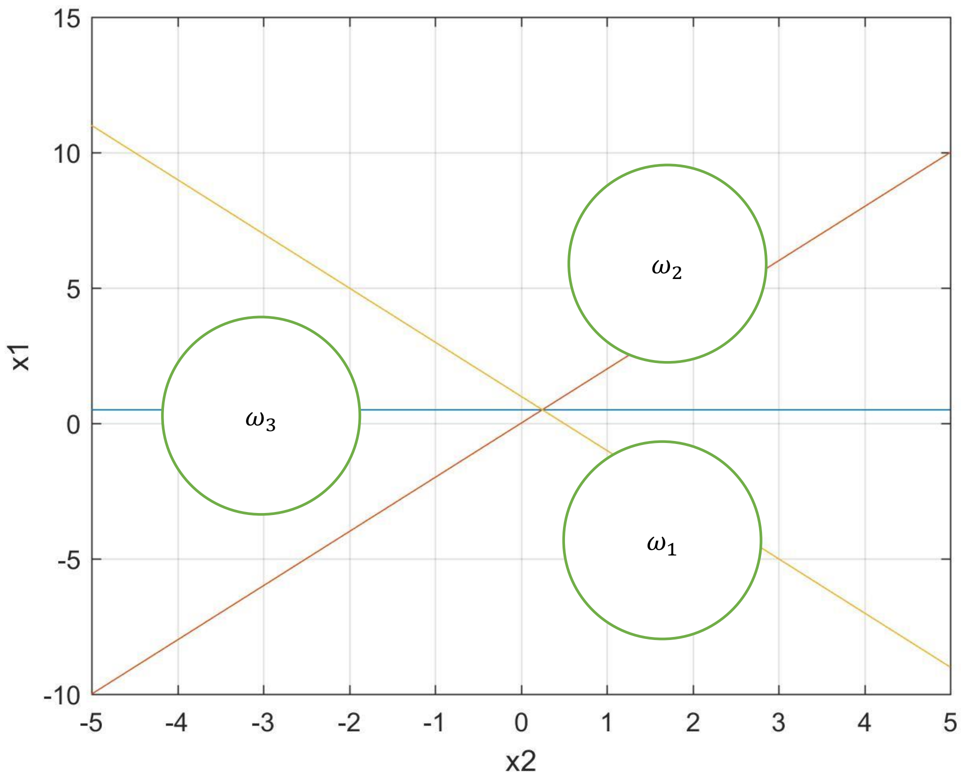
所以，权向量为：

$$a = (0,-3,4)^T$$

2.

根据题意，决策面如下图：

可见，三条决策线交于一点，此时不存在不能判定的区域



3.

（a）

题目可以看做一个约束条件下的非线性规划求解

即：

$$\text{minimize: } \|\mathbf{x}-\mathbf{x}_a\|^2$$
$$\text{constraint: } g(\mathbf{x})=0$$

使用拉格朗日乘子法：

$$\begin{aligned} f(\mathbf{x},\lambda) &= \|\mathbf{x}-\mathbf{x}_a\|^2 + 2\lambda[g(\mathbf{x})] \\ &= \|\mathbf{x}-\mathbf{x}_a\|^2 + 2\lambda[\mathbf{w}^t\mathbf{x}+w_0] \\ &= (\mathbf{x}-\mathbf{x}_a)^t(\mathbf{x}-\mathbf{x}_a)+2\lambda(\mathbf{w}^t\mathbf{x}+w_0) \\ &= \mathbf{x}^t\mathbf{x}-2\mathbf{x}^t\mathbf{x}_a+\mathbf{x}_a^t\mathbf{x}_a+2\lambda(\mathbf{w}^t\mathbf{x}+w_0) \end{aligned}$$

取导数，使其等于 0：

$$\frac{\partial f(\mathbf{x},\lambda)}{\partial \mathbf{x}} = \mathbf{x}-\mathbf{x}_a+\lambda\mathbf{w}=0$$
$$\frac{\partial f(\mathbf{x},\lambda)}{\partial \lambda} = \mathbf{w}^t\mathbf{x}+w_0=0$$

此时向量 x 是 x_a 在超平面上的投影，消去向量 x 解这个方程组，得到：

$$\lambda = \frac{\mathbf{w}^t\mathbf{x}_a+w_0}{\mathbf{w}^t\mathbf{w}}$$

则

$$\|\mathbf{x}-\mathbf{x}_a\| = \left\| \mathbf{x}_a - \left[\frac{\mathbf{w}^t\mathbf{x}_a + w_0}{\mathbf{w}^t\mathbf{w}} \right] \mathbf{w} - \mathbf{x}_a \right\|$$

$$\|\mathbf{x}-\mathbf{x}_a\| = \left\| \left[\frac{\mathbf{w}^t\mathbf{x}_a + w_0}{\mathbf{w}^t\mathbf{w}} \right] \mathbf{w} \right\|$$

$$\|\mathbf{x}-\mathbf{x}_a\| = \left\| \left[\frac{g(\mathbf{x}_a)}{\mathbf{w}^t\mathbf{w}} \right] \mathbf{w} \right\| = \frac{|g(\mathbf{x}_a)|}{\|\mathbf{w}\|}$$

（b）

在（a）中提到向量 x 是 x_a 在超平面中的投影，因此投影为：

$$\mathbf{x} = \mathbf{x}_a - \lambda \mathbf{w}$$

$$= \mathbf{x}_a - \frac{\mathbf{w}^t\mathbf{x}_a + w_0}{\mathbf{w}^t\mathbf{w}} \mathbf{w}$$

$$= \mathbf{x}_a - \frac{g(\mathbf{x}_a)}{\|\mathbf{w}\|^2} \mathbf{w}$$

编程作业：

1（a）

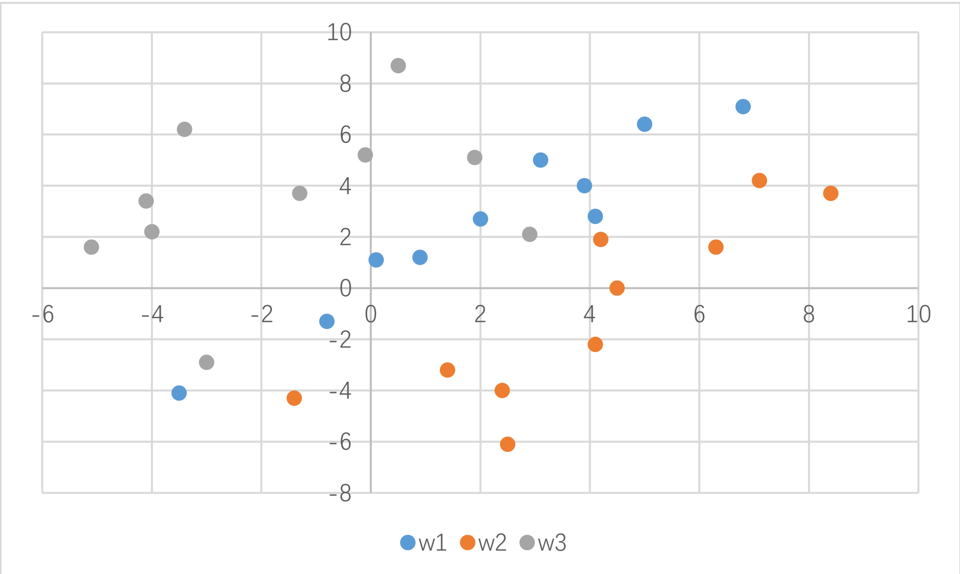
收敛步数为 23 步，结果为 a=[34.0, -30.4, 34.1]

（b）

收敛步数为 16 步，结果为 a=[19.0, -41.4, 48.6]

（c）

见下图：



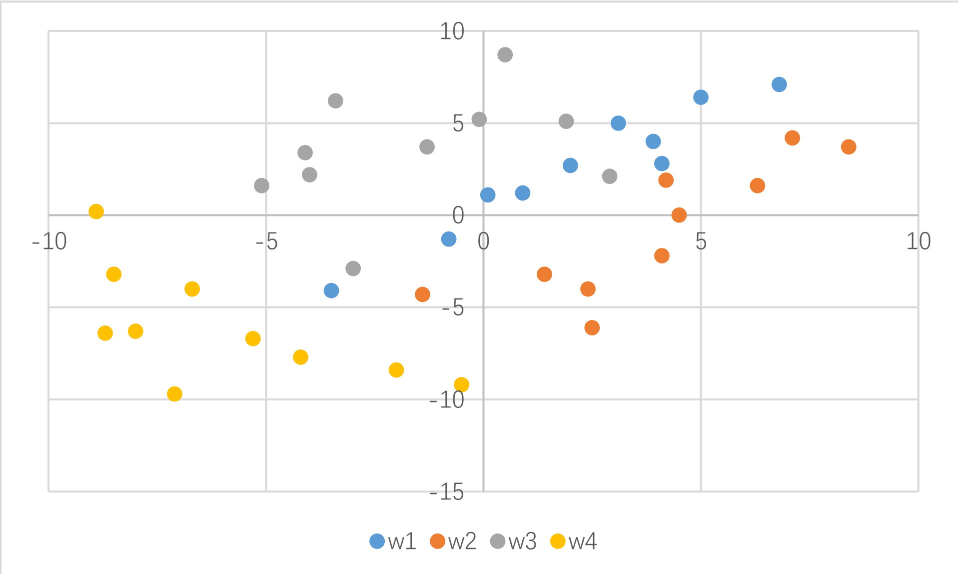
由图中可以看出，w1 和 w2 之间相比 w2 和 w3 和之间，相距更近，更加难以分开，因此迭代次数相对较多

2

训练的参数为：

$$b_{min} = 0.01$$
$$k_{max} = 1000000$$
$$\eta_0 = 0.05$$
$$\mathbf{a} = [0, 0, 0]$$
$$\mathbf{b}_0 = [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]$$

对于 w1 和 w3 的训练数据来说，即使训练次数超过 100 万次，仍然无法拟合，说明了 w1 和 w3 并不是线性可分的（由下图可见。）



而对于 w2 和 w4 来说，当训练到 60006 次时收敛，说明 w2 和 w4 是线性可分的，结果为：

$$b = \begin{pmatrix} 10.25 \\ 1.00 \\ 6.60 \\ 8.46 \\ 7.40 \\ 3.17 \\ 3.34 \\ 2.32 \\ 10.74 \\ 5.24 \\ 1.45 \\ 1.00 \\ 2.35 \\ 2.51 \\ 1.89 \\ 1 \\ 2.47 \\ 4.27 \\ 5.05 \\ 3.82 \end{pmatrix}$$

$$a = \begin{pmatrix} 4.014 \\ 0.575 \\ 0.514 \end{pmatrix}$$